

# Sensitivity analysis of simple global marine biogeochemical models

I. Kriest,<sup>1</sup> A. Oschlies,<sup>1</sup> and S. Khatiwala<sup>2</sup>

Received 18 March 2011; revised 30 March 2012; accepted 25 April 2012; published 5 June 2012.

[1] This study presents results from 46 sensitivity experiments carried out with three structurally simple (2, 3, and 6 biogeochemical state variables, respectively) models of production, export and remineralization of organic phosphorus, coupled to a global ocean circulation model and integrated for 3000 years each. The models' skill is assessed via different misfit functions with respect to the observed global distributions of phosphate and oxygen. Across the different models, the global root-mean square misfit with respect to observed phosphate and oxygen distributions is found to be particularly sensitive to changes in the remineralization length scale, and also to changes in simulated primary production. For this metric, changes in the production and decay of dissolved organic phosphorus as well as in zooplankton parameters are of lesser importance. For a misfit function accounting for the misfit of upper-ocean tracers, however, production parameters and organic phosphorus dynamics play a larger role. Regional misfit patterns are investigated as indicators of potential model deficiencies, such as missing iron limitation, or deficiencies in the sinking and remineralization length scales. In particular, the gradient between phosphate concentrations in the northern North Pacific and the northern North Atlantic is controlled predominantly by the biogeochemical model parameters related to particle flux. For the combined 46 sensitivity experiments performed here, the global misfit to observed oxygen and phosphate distributions shows no clear relation to either simulated global primary or export production for either misfit metric employed. However, a relatively tight relationship that is very similar for the different model of different structural complexity is found between the model-data misfit in oxygen and phosphate distributions to simulated meso- and bathypelagic particle flux. Best agreement with the observed tracer distributions is obtained for simulated particle fluxes that agree most closely with sediment trap data for a nominal depth of about 1000 m, or deeper.

**Citation:** Kriest, I., A. Oschlies, and S. Khatiwala (2012), Sensitivity analysis of simple global marine biogeochemical models, *Global Biogeochem. Cycles*, 26, GB2029, doi:10.1029/2011GB004072.

## 1. Introduction

[2] Coupled global ocean biogeochemical-circulation models serve as a tool to quantitatively assess the role of different biological and biogeochemical processes in the global carbon cycle and their potential sensitivity to environmental changes. In the absence of well-established sets of equations governing the behavior of marine ecosystems, parameterized models of varying structural complexity have been developed. These range from very simple, "nutrient-only" models (mostly based on phosphorus or nitrogen [e.g., Bacastow and Maier-Reimer, 1990]), to models with various autotrophic and heterotrophic compartments, sometimes

describing different elemental cycles [e.g., Aumont and Bopp, 2006; Moore and Doney, 2007]. However, as shown by a recent study [Kriest *et al.*, 2010], an increase in structural complexity (here measured in terms of the number of model components) does not necessarily lead to a better agreement of the model results with observed nutrient and oxygen distributions. Instead, changes in parameter values or parameterizations may be of equal importance for a model's ability to reproduce observed biogeochemical tracer distributions or fluxes, e.g. of CO<sub>2</sub> [Kwon *et al.*, 2009]. Given that structurally very different models may perform similarly in terms of their ability to reproduce observed tracer distributions, the question remains as to whether models of different complexity exhibit a similar robustness in terms of biogeochemical tracer fluxes and their sensitivity to environmental change.

[3] In a pilot study [Kriest *et al.*, 2010], we considered a hierarchy of models that differed in both structural complexity and in various process parameterizations. While our goal there was to investigate the impact of model complexity on model performance, parameter values employed in that study were mostly chosen subjectively according to "best" estimates. Model-data misfits could, therefore, potentially

<sup>1</sup>GEOMAR, Helmholtz Centre for Ocean Research Kiel, Kiel, Germany.

<sup>2</sup>Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, USA.

Corresponding author: I. Kriest, GEOMAR, Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, DE-24105 Kiel, Germany. (ikriest@ifm-geomar.de)

Copyright 2012 by the American Geophysical Union.  
0886-6236/12/2011GB004072

**Table 1.** Biogeochemical Parameters, Their Meaning and Units for the Different Sensitivity Experiments<sup>a</sup>

Parameter	Meaning	Unit
<i>Phytoplankton</i>		
$\mu_{PHY}^b$	max. growth rate	$d^{-1}$
$I_c$	half-sat. constant for light	$W\ m^{-2}$
$K_{PHY}$	half-sat. constant for $PO_4$ uptake	$mmol\ P\ m^{-3}$
<i>Dissolved Organic Matter, DOP (N-DOP, NPZD-DOP)</i>		
$\lambda'_{DOP}$	decay rate (all layers)	$yr^{-1}$
$\sigma_{DOP}$	fraction production assigned to DOP	
<i>Zooplankton (NPZD-DOP Only)</i>		
$\mu_{ZOO}$	max. growth rate	$d^{-1}$
$K_{ZOO}$	half-sat. constant for $PO_4$ uptake	$mmol\ P\ m^{-3}$
$\kappa_{ZOO}$	quadratic mortality	$(mmol\ P\ m^{-3})^{-1}d^{-1}$
<i>Detritus (NPZD-DOP Only)</i>		
$\lambda'_{DET}$	decay rate	$d^{-1}$
$a$	increase of sinking speed with depth	$d^{-1}$
<i>Particle Flux (N, N-DOP Only)</i>		
$r/a$	exponent of power law flux	

<sup>a</sup>See Table 2 for parameter values of set 1, and Table 3 for parameters values of set 2.

<sup>b</sup>For N and N-DOP  $\mu_{PHY}$  refers to an assumed, constant phytoplankton concentration of  $0.0028\ mmol\ P\ m^{-3}$  (see text).

arise from an unintended selection of some particularly inappropriate model parameters. In order to rule out this possibility, and thus more fully assess a particular model's skill and sensitivity, we have carried out a "coarse sweep" of the parameter space of three relatively simple global biogeochemical models. By restricting ourselves to model structures with a maximum of up to six tracers ( $PO_4$ , dissolved organic phosphorus, phyto- and zooplankton biomass, detritus, and  $O_2$ ) we keep the number of parameters to be investigated manageably small. Moreover, with this relatively small number of prognostic tracers and an efficient method for physical tracer transport, the "transport matrix method" [Khatiwala *et al.*, 2005], we are able to spin the models up to equilibrium starting from globally homogeneous tracer concentrations within a reasonable amount of time. Neglecting exchange via rivers or the seafloor, total phosphorus is conserved. In the absence of strong biogeochemical non-linearities that could lead to bifurcations and multiple equilibria, the resulting seasonally cycling steady state tracer distributions should be independent of initial conditions. This was confirmed in a sensitivity experiment in which different initial distributions of phosphorus and oxygen resulted in virtually identical distributions after 3000 years. For a given phosphorus inventory, the spun-up state should therefore reflect only the combined effects of the underlying circulation field (which here is identical for all simulations) and the representation of biogeochemistry (which is not identical among the simulations). Differences in the different models' fits to global data sets of  $PO_4$  and  $O_2$  can thus be generated only by differences among the different biogeochemical model formulations.

[4] By comparing the spun-up model solutions with observed tracer distributions we aim at addressing the following questions: (1) How sensitive are the model results to variations in the biogeochemical model parameters compared to variations in biological model structure?; (2) Which

regions of the World Ocean are most sensitive to changes in the model parameters?; (3) To what extent can other observations, such as global primary production, export production or deep particle flux, help to constrain a model's ability to reproduce observed global phosphate and oxygen distributions?; and (4) Is there a single "best" model that satisfies several different misfit metrics?

[5] This paper is organized as follows. After a brief description of the various models, we present an assessment of the models' performances relative to the tracer distribution in the entire water column of the global ocean. This global assessment employs different metrics and is supplemented by an investigation of the model-data misfits in the surface layer, which concentrates on the more dynamic biological processes and associated shorter timescales. We then examine simulated biogeochemical tracer distributions in different oceanic regions for their sensitivity to changes in model parameters. Lastly, we analyze the relation of different metrics of model-data tracer misfits to simulated global fluxes such as primary production, export production, and particle flux.

## 2. Model and Experiment Description

### 2.1. The Tracer Transport Model

[6] Carrying out many biogeochemical tracer simulations that are to be integrated into a seasonally cycling equilibrium requires a computationally efficient tracer transport routine. In this study we use the "transport matrix method" (TMM) of Khatiwala *et al.* [2005], an efficient offline method for the simulation of biogeochemical tracers in the ocean. As in Kriest *et al.* [2010] (hereinafter KKO), we employ transport dynamics derived from a  $2.8^\circ$  global configuration of the MIT ocean model [Marshall *et al.*, 1997] with 15 vertical levels that had been driven by seasonally cycling climatological surface fluxes of momentum, heat, and freshwater. An extensive description of the TMM can be found in Khatiwala *et al.* [2005] and Khatiwala [2007].

### 2.2. The Biogeochemical Models

[7] The simplest biogeochemical model, labeled "N" for nutrient, used in this study considers only  $PO_4$  as nutrient, similar to the model of Bacastow and Maier-Reimer [1990]. A modification to this model additionally includes dissolved organic phosphorus (DOP). This yields the second model, "N-DOP", which is similar to the models used by Bacastow and Maier-Reimer [1991] and Parekh *et al.* [2005]. The third and more complex model we use, "NPZD-DOP", is similar to that of Schmittner *et al.* [2005]. All models are described in detail in KKO. Model equations can be found in Appendix A, and a description of model parameters is given in Table 1. For each of the three models, we focus on parameter sensitivity experiments with respect to the reference simulations described by KKO, which are here denoted as N-ref, N-DOP-ref and NPZD-DOP-ref.

### 2.3. Parameter Sensitivity Experiments

[8] Two sets of sensitivity experiments are carried out. The first set explores changes in parameters related to local processes (e.g., phytoplankton and zooplankton growth and loss terms, DOP production and decay terms; see Table 2), which do not directly affect the redistribution of nutrients

**Table 2.** Set 1 of Parameter Sensitivity Experiments With Regard to Autotrophic and Heterotrophic Processes of the Different Biogeochemical Models<sup>a</sup>

	Experiment																	
	<i>ref</i> <sup>b</sup>	<i>p1</i>	<i>p2</i>	<i>p3</i>	<i>p4</i>	<i>p5</i>	<i>p6</i>	<i>d1</i>	<i>d2</i>	<i>d3</i>	<i>d4</i>	<i>d5</i>	<i>z1</i>	<i>z2</i>	<i>z3</i>	<i>z</i>	<i>z5</i>	<i>z6</i>
<i>N</i>																		
$\mu_{PHY}$ <sup>c</sup>	0.66	0.44	2.0															
$I_c$	30			15	45													
$K_{PHY}$	0.5					0.25	1.0											
<i>N+DOP</i>																		
$\mu_{PHY}$ <sup>c</sup>	2.0	0.66	6.0															
$I_c$	30			15	45													
$K_{PHY}$	0.5					0.25	1.0											
$\lambda'_{DOP}$	0.5							0.25	1.0	0.25	0.50	1.0						
$\sigma_{DOP}$	0.67									0.33	0.33	0.33						
<i>NPZD+DOP</i>																		
$\mu_{PHY}$	0.6	0.3	1.2															
$I_c$	24			12	48													
$K_{PHY}$	0.031					0.016	0.063											
$\lambda'_{DOP}$	0.17							0.5	2.0									
$\mu_{ZOO}$	2.0												1.0	4.0				
$K_{ZOO}$	0.088														0.063	0.125		
$\kappa_{ZOO}$	3.2																1.6	6.4

<sup>a</sup>See Table 1 for meaning and units of the different parameters.<sup>b</sup>These values have been used for the standard experiments (N-ref, N-DOP-ref and NPZD-DOP-ref) and the experiments of set 2.<sup>c</sup>For N and N-DOP  $\mu_{PHY}$  refers to an assumed, constant phytoplankton concentration of 0.0028 mmol P m<sup>-3</sup>.

and oxygen in the vertical. These experiments are denoted by numbers *p1*-*p6* for parameters related to phytoplankton growth (models N, N-DOP and NPZD-DOP), by numbers *d1*-*d5* for parameters related to DOP production and decay (models N-DOP and NPZD-DOP), and by numbers *z1*-*z6* for parameters related to zooplankton physiology (model NPZD-DOP only). For all experiments in this set, we assume a power law profile for the particle flux function (models N and N-DOP) or vertically increasing sinking speed of detritus (NPZD-DOP), as explained below in more detail. The second set of experiments investigates the impact of different parameters for particle flux and remineralization. These experiments are denoted by *s1* to *s4*.

### 2.3.1. Set 1: Biological Parameters

[9] In experiments *p1* and *p2* of all models we varied the maximum production rate,  $\mu_{PHY}$  (note that in model N this rate represents the export production rate). In experiments *p3* to *p6* we varied half-saturation constants for nutrient and light limitation ( $K_{PHY}$  and  $I_c$ , respectively). Experiments *d1* and *d2* of models N-DOP and NPZD-DOP are related to the DOP decay rate,  $\lambda'_{DOP}$ . For model N-DOP, in experiments *d3* to *d5* we additionally varied the DOP decay rate on the background of a lower production rate of DOP,  $\sigma_{DOP}$ . Experiments *z1* to *z6* carried out with model NPZD-DOP are related to zooplankton growth and loss parameters ( $\mu_{ZOO}$ ,  $K_{ZOO}$ , and  $\kappa_{ZOO}$ ). The parameter values of the respective experiments are displayed in Table 2. Typically, sensitivity runs are performed with parameters increased or decreased by a factor two with respect to the parameter settings of the respective reference model.

### 2.3.2. Set 2: Particle Flux Profiles

[10] For each of the model structures, we further carried out four experiments related to changes in the particle flux function. Since model NPZD-DOP differs from N and N-DOP in that it explicitly resolves sinking detritus, in the following we describe the different flux parameterizations

in the light of the underlying, implicit assumptions of particulate organic matter (POM) settling speed [see also *Kriest and Oeschlies*, 2008].

[11] Assume that the settling speed of POM,  $w$ , can be described by a one-parameter function of depth  $z$ :

$$w = az, \quad (1)$$

and that POM remineralizes with a constant remineralization rate,  $r$ . With  $a > 0$ , in equilibrium, and in the absence of significant current velocities or numerical effects, we would recover a power law,  $F \propto z^{-(r/a)}$ , for the downward particle flux.

[12] In the experiments of set 1 (experiments *ref*, *p1*-*p6* and *d1*-*d5* of N and N-DOP), we adopt the value of *Martin et al.* [1987] of  $r/a = 0.858$ . The value for  $r/a$  is changed by  $\pm 25\%$  in experiments *s2* and *s3*, and by  $\pm 50\%$  in experiments *s1* and *s4*. Using this range of exponents we basically follow the approach by *Kwon et al.* [2009] and approximately cover the range consistent with sediment trap data [*Martin et al.*, 1987; *Buesseler et al.*, 2007].

[13] Model NPZD-DOP includes an explicit detritus component, sinking with a settling speed given by equation (1), and remineralizing at a constant rate  $\lambda'_{DET}$  of 0.05 d<sup>-1</sup> (corresponding to  $r$  of the direct flux formulation used for N and N-DOP). As noted above, assuming equilibrium conditions, with  $a$  of equation (1) set to 0.0583 d<sup>-1</sup> the long-term particle flux profiles of model NPZD-DOP-ref should theoretically correspond to those of set 1 models N and N-DOP. As for the two simpler models, we carried out four NPZD-DOP sensitivity experiments where the rate of vertical increase of settling speed  $a$  is changed such that  $r/a = \lambda'_{DET}/a$  changes by  $\pm 25\%$  in experiments *s2* and *s3*, and by  $\pm 50\%$  in experiments *s1* and *s4* (see also Table 3).

[14] Export production is simulated as an immediate loss of a fraction of production (given by  $1 - \sigma_{DOP}$ ) from each of

**Table 3.** Set 2 of Parameter Experiments With Regard to Sinking Speed and Remineralization of the Different Biogeochemical Models<sup>a</sup>

	Experiment				
	s1	s2	ref <sup>b</sup>	s3	s4
$r/a$	1.287	1.073	0.858	0.644	0.429

<sup>a</sup>Biological parameters of all experiments of set 2 are the same as in N-ref, N-DOP-ref and NPZD-DOP-ref (see Table 2). All experiments assume a detritus decay rate  $\lambda_{DET} = r = 0.05 \text{ d}^{-1}$ . See text for further details, and Table 1 for meaning and units of the different parameters.

<sup>b</sup>This value has been used for the standard experiments (N-ref, N-DOP-ref and NPZD-DOP-ref) and the experiments of set 1.

the upper two layers (see KKO for a detailed description) in models N and N-DOP. In model NPZD-DOP, the export production is calculated as the flux of sinking detritus out of layer 2 (i.e., across  $z = 120 \text{ m}$ ).

[15] In the figures discussed below, we denote the respective reference experiment by a large open circle, and the experiments related to parameter set 1 by small circles and/or lines. Experiments related to the sinking parameters (set 2) are denoted by inverted triangles and/or lines.

### 3. Model Assessment

[16] In all experiments, the model was spun up from globally uniform values of  $\text{PO}_4$  and  $\text{O}_2$  for 3000 years. Model-data misfits were evaluated for simulated annual means of phosphate and oxygen of year 3001 of the respective model runs. As a first step toward quantifying the model-data misfit, we use the global root mean square concentration error,  $\sqrt{G}$ :

$$G = \overline{(\mathcal{M} - \mathcal{O})^2} \quad (2)$$

between model  $\mathcal{M}$  and observation  $\mathcal{O}$ . Models are assessed with respect to annual mean  $\text{PO}_4$  and  $\text{O}_2$ , as provided by the World Ocean Atlas 2005 (WOA) [Garcia *et al.*, 2006a, 2006b]. We use the non-interpolated observations, to avoid any potential biases due to interpolation (although, for typical model-data misfits, these tend to be small; see KKO). Details regarding the treatment of data and regridding procedure can be found in KKO.

[17] Besides the total misfit we also investigate the misfit between model and observations restricted to the surface layer (0–50 m),  $G_S$ . Because of the rapidity of air-sea gas exchange, surface  $\text{O}_2$  is relatively insensitive to biological parameter variations. Therefore  $G_S$  evaluates only the misfit with respect to annual mean surface  $\text{PO}_4$ . Further, by calculating the misfit for specific regions ( $R$ ) we examine how well a model represents the tracer distribution in this region. Again, by restricting the regional misfit to the upper 0–50 m, we investigate how well the model represents processes at the surface ( $R_S$ ).

### 4. Model Sensitivity to Parameter Variations

[18] The sensitivity of the model-data misfits to the choice of the model parameter values is, for each model, investigated by a “coarse sweep” of the parameter space. As noted above, each parameter of the respective model structure’s

standard configuration is varied by typically half and twice the standard value for the local parameters (Table 2), and by  $\pm 50\%$  and  $\pm 25\%$  for the sinking parameters (Table 3). We recognize that this approach does not, by any means, result in a complete scan of the parameter space. Instead, it is a pragmatic and computationally feasible compromise that, we believe, yields useful information about the sensitivity of the model-data misfit, and thereby our measure of model quality, to the particular parameter choices. Our eventual goal is to perform a full scan of parameter space by exploiting computationally even more efficient methods for integrating the individual model runs into steady state [e.g., Li and Primeau, 2008; Khatiwala, 2008].

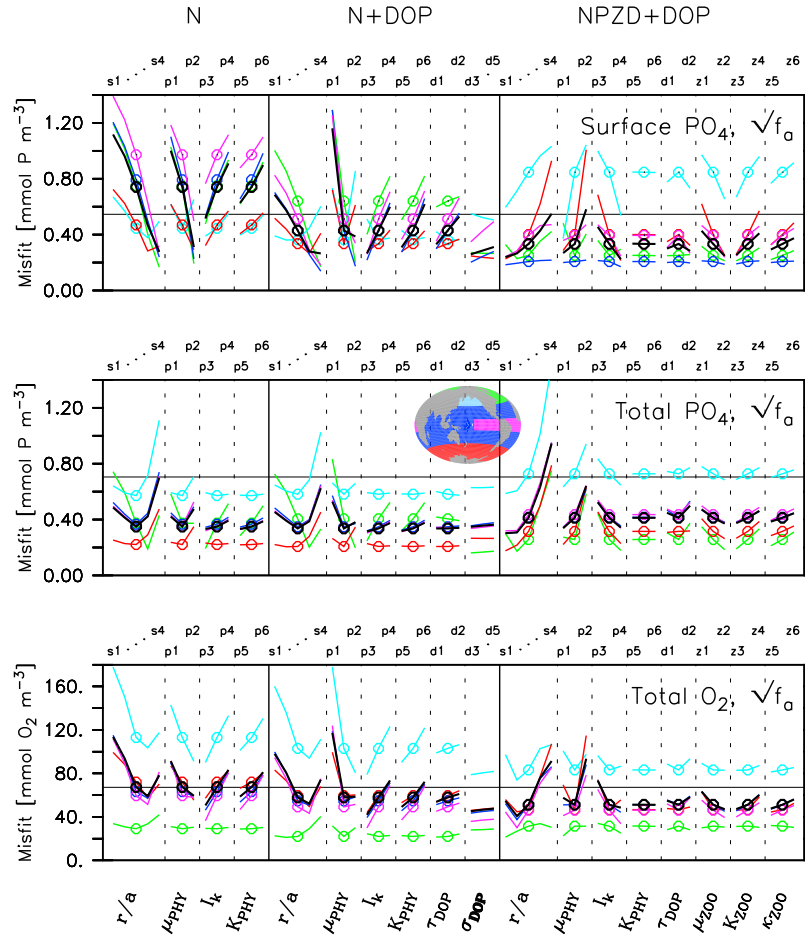
#### 4.1. Global Misfit to Surface Phosphate

[19] Considering the misfit to observed surface  $\text{PO}_4$  in the surface layer ( $G_S$ ), the sensitivity to variations of individual parameters is generally higher for the simpler models N and N-DOP than for model NPZD-DOP (Figure 1, top).

[20] The surface  $\text{PO}_4$  misfit shows a very pronounced sensitivity of the structurally simple N and N-DOP models to the particle flux parameters, and additionally also to the production parameters. The misfit decreases with increasing sinking velocity for the N and N-DOP models, whereas it decreases with decreasing sinking velocity for the NPZD-DOP model. In the N and N-DOP models, deeper penetration of the particle flux into the ocean, as mediated via a low power law exponent reduces the surface misfit, because it reduces the overall surface nutrients (see also Figure 2). For the same reason, any increase in production, as mediated via an increase in  $\mu_{PHY}$ , or via a decrease in the half-saturation constant for light or nutrients ( $I_c$  and  $K_{PHY}$ , respectively), also reduces the misfit to observed surface phosphate for these models. In contrast, the more complex model NPZD-DOP shows a smaller variation of the surface misfit with respect to variations of parameters controlling biological turnover in the euphotic zone.

[21] One reason for the different sensitivities of structurally simple versus complex models with respect to the surface  $\text{PO}_4$  misfit can be found in the generally too high surface nutrients in the reference configuration of the simple models (see Figure 2), caused by rather low growth rate and high half-saturation constants in the standard run. An enhanced particle flux or a higher production rate can substantially reduce surface phosphate concentrations (see also Figure 2), thus improving the global fit to data.

[22] In contrast to the structurally simple models, the reference run of model NPZD-DOP already has quite high production parameters, and therefore relatively low nutrients in the surface layers (Figure 2) which, in some places, lead to substantial underestimates of the observations. As a consequence, reductions in growth and/or export improve surface phosphate concentrations in this model, as do parameters that induce a stronger top-down control via zooplankton grazing (Figure 1). The beneficial effect of parameters that induce slow phytoplankton production is most pronounced in so-called HNLC (high-nutrient-low-chlorophyll) regions (e.g., the Southern Ocean, equatorial regions or the northern North Pacific), which may reflect a compensation for the lack of an explicit iron limitation in the model (see section 5).



**Figure 1.** Different global misfit functions for models N, N-DOP and NPZD-DOP (black lines). (top) Misfit to surface  $\text{PO}_4$ , (middle) misfit to  $\text{PO}_4$ , and (bottom) misfit to  $\text{O}_2$  (see equation (2) and section 3). For each parameter, the experiment with the smallest parameter value is plotted first, then the value increases to the right, with the reference run plotted as open circle for comparison in the middle. Colored lines denote the model's fit in different regions. Green: northern North Atlantic. Light blue: northern North Pacific. Dark blue: low latitudes, except equatorial upwelling. Red: Southern Ocean. Magenta: equatorial upwelling region. See also shaded area in inset of Figure 1 (middle) for location and extent of regions. Horizontal thin line denotes the global standard deviation of the observations.

#### 4.2. Global Misfit to Phosphate and Oxygen

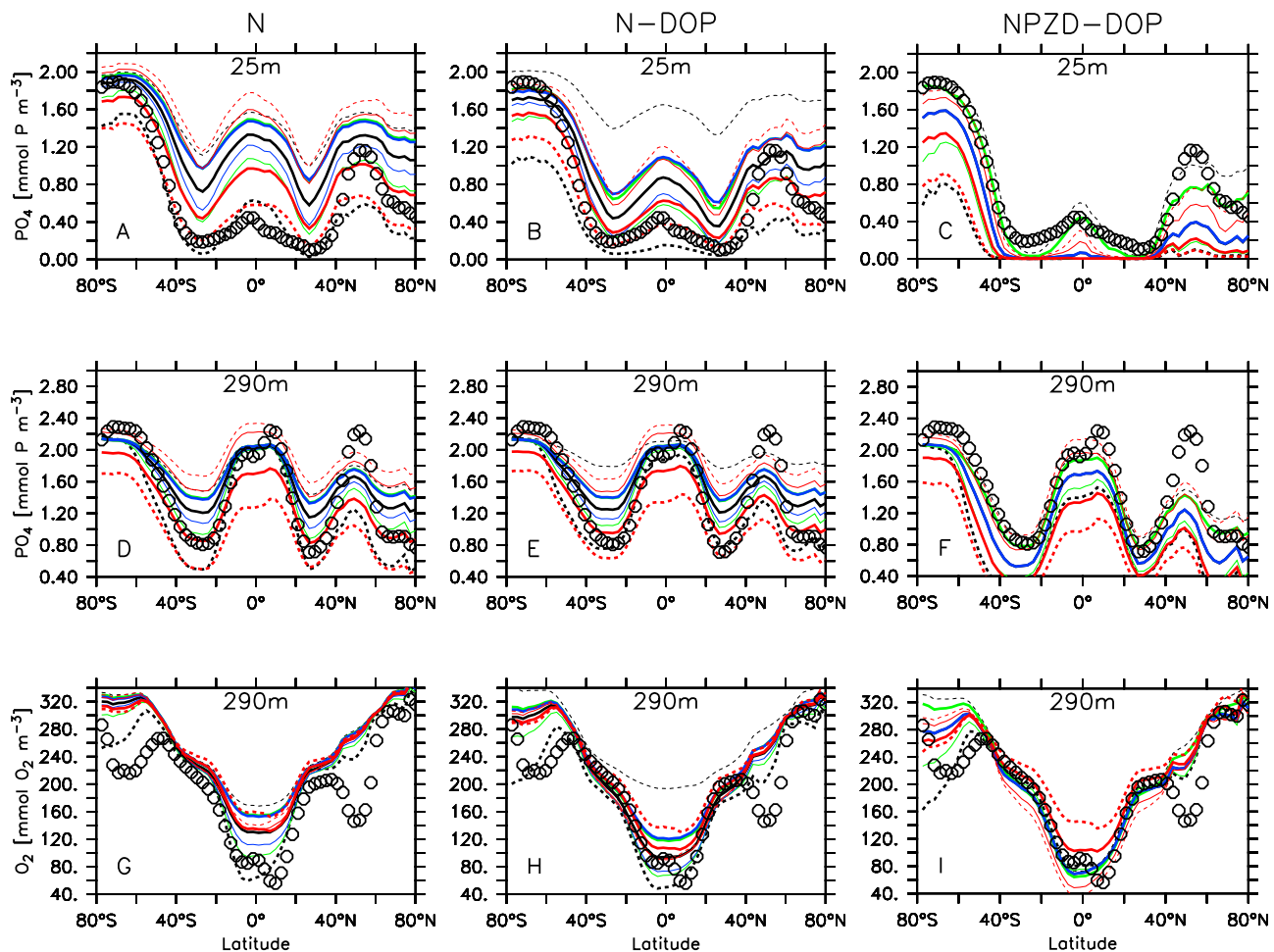
[23] For the different metric considered, the global misfits to observed  $\text{PO}_4$  and  $\text{O}_2$  distributions tend to vary more with parameter values than with model type (Figure 1). According to our metrics, all models seem to be particularly sensitive to particle flux or sinking parameters (i.e., experiments  $s1$ – $s4$ ). Based on the misfit function to the global distribution of  $\text{PO}_4$  (Figure 1, middle), a small power law flux exponent (i.e., deeper penetration of flux into the ocean interior) or, equivalently, sinking speed increasing rapidly with depth, as, for example, in experiment  $s4$ , is detrimental for all models. One reason for this degradation are the too low phosphate concentrations in the mesopelagic zone caused by the excessively large transport of organic matter to the ocean interior in exp.  $s4$  (thick dashed red lines in Figure 2).

[24] The structurally simple models N and N-DOP show best fits to observed phosphate with the standard power law exponent, while the more complex model NPZD-DOP gives best results with a sinking speed increasing only slowly with

depth (NPZD-DOP, experiment  $s2$ ). The simple models show relatively little response to changes in growth, production or DOP-related parameters, but model NPZD-DOP can be slightly improved by a decrease in growth rate, as mediated via a decrease in maximum growth rate ( $\mu_{PHY}$ , experiment  $p1$ ) or light affinity (increase in  $I_c$ , experiment  $p4$ ).

[25] With respect to the reference experiment, the fit of models N and N-DOP to the observed oxygen distribution (Figure 1, bottom) can be improved by slightly decreasing the power law exponent (experiments  $s3$ ), or by increasing the simulated biological production via a high light affinity in experiments  $p3$ . As for phosphate, a too strong particle flux (as in experiments  $s4$  of N and N-DOP) reduces the fit. The NPZD-DOP model is about as sensitive to the parameterization of particle flux as are the simpler models. However, its fit to the global oxygen distribution is best for a relatively small sinking velocity of detritus (experiment  $s2$ ).

[26] For all misfit functions discussed so far, the largest contribution to the global misfit for almost all models and experiments comes from the low latitudes, most likely owing



**Figure 2.** Zonal average of  $\text{PO}_4$  for models (a and d) N, (b and e) N-DOP, and (c and f) NPZD-DOP, at a depth of 25 m (Figures 2a–2c) and 290 m (Figures 2d–2f). (g–i) Zonal average of  $\text{O}_2$ , 290 m. Lines indicate different model experiments. Thick black: exp. *ref.* Thin and thick dashed black: exps. *p1* and *p2*, respectively. Thin and thick green: exps. *p3* and *p4*, respectively. Thin and thick blue: exps. *p5* and *p6*, respectively. Thin and thick red: exps. *s2* and *s3*, respectively. Thin and thick dashed red: exps. *s1* and *s4*, respectively. Open circles denote observations (interpolated data set of Garcia *et al.* [2006b, 2006a]). Note that in the plot for NPZD-DOP, black and thin blue lines are hidden under the thick blue line.

to their large spatial extent. Although volume specific misfits are generally largest in the North Pacific (cyan lines in Figure 1, see discussion below), because of the small size of this region it contributes only little to the global misfit.

## 5. Model Skill in Different Regions

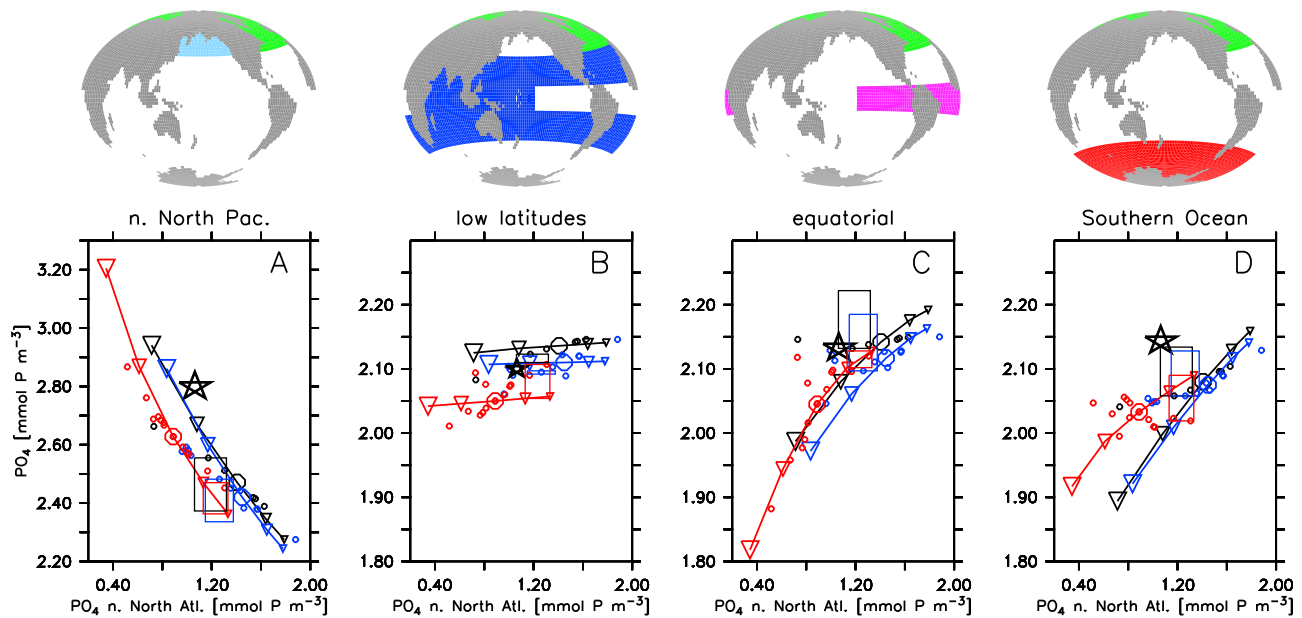
### 5.1. Regional Misfit to Surface Phosphate

[27] The surface misfit averaged over various domains ( $R_S$ ) is shown in Figure 1 (top). We find that for the parameter combinations explored, the simple models perform particularly poorly at the surface in the equatorial upwelling region, in the low latitudes, and in the northern North Atlantic, mostly due to an overestimate of surface nutrients (see also Figure 2). NPZD-DOP, in contrast, shows severe deficiencies in the northern North Pacific, and also in the Southern Ocean, which can be attributed to a substantial underestimate of surface nutrients (see also Figure 2). While the simple models' misfit may indicate a too weak biological

turnover, the fast growth rates used in the reference configuration of model NPZD-DOP may not be appropriate for HNLC regions.

[28] It is important to note that none of the models used here explicitly account for iron limitation. The above result might therefore be interpreted as a sign of missing iron limitation in the models: while the low- to medium-growth cases of the simple models, or NPZD-DOP with low growth, mimic iron limited growth in the northern North Pacific and Southern Ocean on the expense of simulating low growth also in the North Atlantic, the fast-growing NPZD-DOP scenarios mimic non-limited growth in the northern North Atlantic on the expense of simulating too fast growth in the North Pacific and Southern Ocean. However, it is still possible that the inability of the various models to accurately simulate surface concentrations in both regions simultaneously is related to the particular combinations of parameters we have thus far explored. At this stage we cannot rule out that a more comprehensive search of parameter





**Figure 3.** Average phosphate in different regions versus phosphate in the northern North Atlantic (nNA). (a) Northern North Pacific versus nNA. (b) Low latitudes versus nNA. (c) Equatorial upwelling versus nNA. (d) Southern Ocean versus nNA. Color denotes model type (black: N; blue: N-DOP; red: NPZD-DOP). Symbol type denotes model experiments: Large open circle: reference scenario; small circles: experiments of set 1. Inverted triangles connected with bold lines: experiments  $s1$  to  $s4$  related to particle flux (set 2). Note that the size of the triangles increases with increasing sinking speed, i.e., from experiments  $s1$  to  $s4$ . Rectangles denote the range of concentrations for the best three experiments of each model. Observations are shown by black open stars.

space or model optimization might allow to accurately simulate both regions even without an explicit consideration of iron limitation. A follow-on study is planned to specifically investigate the extent to which an explicit inclusion of iron limitation can improve the model fit.

## 5.2. Regional Misfit to Phosphate and Oxygen

[29] All models show highest deficiencies with respect to the total (surface + deep) phosphate and oxygen misfit in the northern North Pacific (Figures 1 (middle) and 1 (bottom)). For this region, the volume-average regional misfit ( $R$ ) of both phosphate and oxygen is always near or greater than the global variance of the annually averaged observed tracer distribution (horizontal black lines in Figure 1). Given the large age of North Pacific deep waters [Khaliwala *et al.*, 2012], this area is a likely candidate to accumulate deficiencies of both the biological and physical model, for example associated with parameters related to sinking and remineralization. Variations in parameters related to particle flux exert, in our experiments, the largest influence on the misfit in the North Pacific. Thus, although the surface misfit suggests that the structurally simple models may be able to reproduce surface properties in this region better than the more complex models, this is not reflected at depth, where the older waters have also accumulated signals of remote biogeochemical processes.

## 5.3. Northern North Pacific

[30] The high sensitivity of simulated nutrient concentrations in the northern North Pacific to changes in parameter values is also evident from Figure 3, which plots average

phosphate concentrations in the different oceanic regions (northern North Pacific, low latitudes, equatorial regions and the Southern Ocean) versus those of the northern North Atlantic. The different models respond to an increase in the remineralization length scale with a decrease of phosphate in the northern North Atlantic, and a corresponding decrease in the Southern Ocean and the equatorial upwelling region. Phosphate concentrations in the subtropical latitudes are largely unaffected by an increase in sinking speed. In contrast, average phosphate in the northern North Pacific is inversely related to the average phosphate in the northern North Atlantic. The “faster” the detritus export is in a model (as, for example, reflected in a low power law exponent), the more phosphate is transported by the global thermohaline “conveyor belt” circulation into the northern North Pacific. The trend of inversely related phosphate concentrations also shows up in the other parameter experiments.

[31] Interestingly, the models with the smallest global misfit (indicated by rectangles in Figure 3a) cluster around a ratio of  $\approx 2.4 \text{ mmol P m}^{-3}$  (North Pacific) to  $\approx 1.2 \text{ mmol P m}^{-3}$  (North Atlantic) which differs from the observed ratio of  $\approx 2.8:1$  (star in Figure 3a). That is, the apparent globally best fit of these models to observed phosphate is achieved via an underestimate of the North Pacific’s average phosphate content and a slight overestimate of the North Atlantic’s average phosphate content.

[32] One reason for this apparent discrepancy between the fit to bulk regional phosphate and global misfit function is the small contribution of the North Pacific to the global misfit. Another reason for the apparent discrepancy can be found in the detailed spatial resolution of the misfit function:

A relatively “good” match of regionally averaged phosphate of the models with fast sinking speed is achieved via a local underestimate of surface concentrations together with a local overestimate of deep concentrations. These misfits of opposite sign result in a higher misfit for this region, but in a matching regional average.

## 6. Discussion

[33] Our results suggest that the particle flux and/or sinking parameters are particularly important for an accurate description of the global distribution of phosphate and oxygen. This holds for all three structurally different biogeochemical models considered. The importance of the particle flux parameter for distributing phosphate among the northern North Pacific and northern North Atlantic has also been found by *Kwon and Primeau* [2006] in a sensitivity analysis of a structurally simple model. They found a net export of phosphate from the North Atlantic to the North Pacific when the power law exponent was decreased (i.e., an increase in the strength of the biological carbon pump). Likewise, *Bacastow and Maier-Reimer* [1991] found an increase of phosphate in the meso- and bathypelagial of the northwestern Pacific upon increasing the length scale of their simulated exponential particle flux (compare *Bacastow and Maier-Reimer* [1991, Figures 5b and 5c]). The results presented here qualitatively agree with these earlier studies.

[34] The misfit of the structurally simple models to phosphate suggests that the optimum power law exponent of the simple models - at least when simulated with the other parameters set to their default values - is close to the open-ocean composite of 0.858 suggested by *Martin et al.* [1987]. Values near 0.9 have also been widely used by the global modeling community. A somewhat higher optimum value of 1.0 was found by *Kwon and Primeau* [2006]. The more complex NPZD-DOP model reaches smallest global misfits to both phosphate and oxygen with a sinking speed corresponding to a power law exponent of 1.07. The fact that the more complex model seems to favor a somewhat lower sinking velocity might be explained by a more efficient transfer of surface nutrients into particulate organic matter, part of which can be exported to deeper layers by vertical mixing. Additionally, the effect of numerical diffusion caused by the upstream scheme for particle sinking can be comparable to a  $\approx 12\%$  decrease in the power law exponent [*Kriest and Oschlies*, 2011]. Accounting for this effect would probably shift the “best” estimate of the power law exponent for model NPZD-DOP toward a value of  $\approx 1$ .

### 6.1. How Robust Is the Identified Misfit Minimum?

[35] Despite the fact that the derived “optimum” flux exponent is near values found by other studies, the more complex model may be especially prone to so-called “local” misfit minima, i.e., the occurrence of many different optimum values in (a potentially high-dimensional) parameter space. (For the NPZD-DOP model as examined here, for example, this is an 8-dimensional space.) That is, given a different default set of parameters, the optimum flux exponent may be higher or lower. Only a complete and finely resolved examination of the parameter space could provide a conclusive answer about the optimum parameter set. This would require many evaluations of the spun-up model system.

While this can be attempted for a structurally simple model, [e.g., *Kwon and Primeau*, 2006, 2008], this computationally very demanding task has, to our knowledge, not yet been carried out for more complex models.

[36] To get a first impression about the robustness of our results, we have carried out an additional set of experiments with model NPZD-DOP. This supplements the above experiments in the following way: (1) Two more evaluations of the model are carried out in the vicinity of the “Martin” flux exponents, namely by varying  $a$  of equation (1) such that  $r/a = 0.751$  and  $r/a = 0.965$ , i.e. deviating  $\pm 12.5\%$  from the standard value. (2) The extended set of flux-exponent sensitivity experiments is repeated with a half-saturation constant for light of  $I_c = 48$ , i.e., twice the default value. As noted above and shown in Figure 1, this value - with all other parameters set to default values - improves the model fit of NPZD-DOP. This latter set of experiments therefore also examines the question whether we can further improve the model solution by combining two “beneficial” parameter choices at the same time.

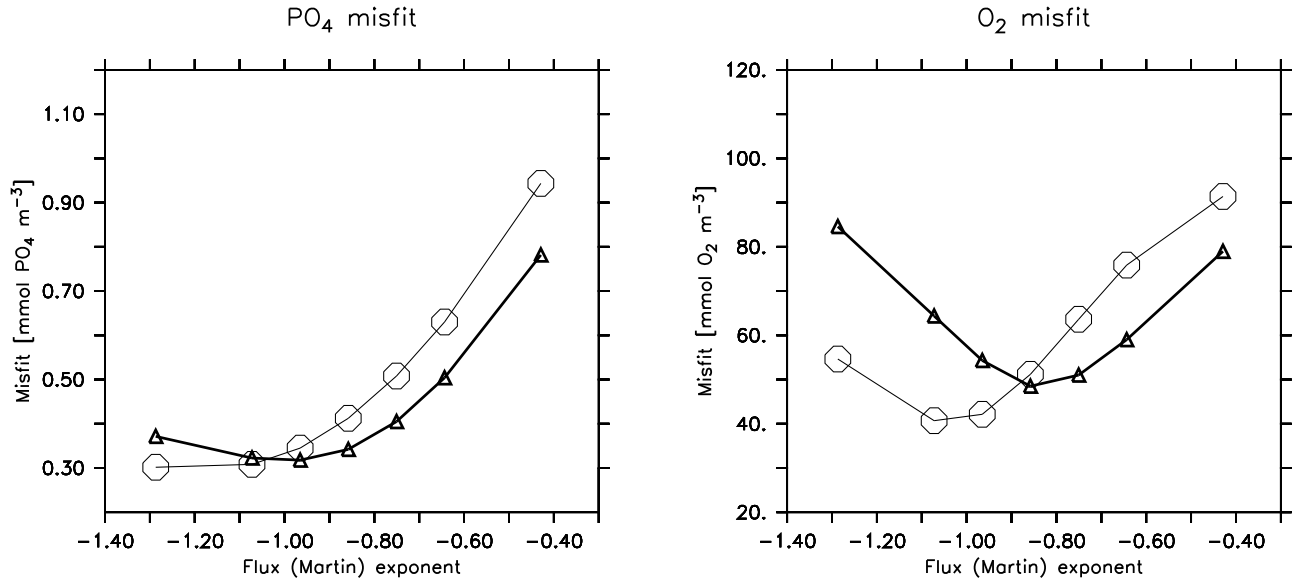
[37] Unfortunately, Figure 4 shows that this is not the case: the misfit to observations remains roughly the same over the range of export exponents tested (phosphate), or even deteriorates (oxygen). For both tracers the optimum flux exponent decreases when the half-saturation constant for light is doubled. That is, the decrease in light sensitivity is compensated by a trend toward faster sinking speed. The finer resolution of the parameter space around the default value of 0.858 reveals no discontinuities. The apparent smoothness of the misfit function for both values of  $I_c$  suggests that relatively simple search algorithms may be able to better constrain the optimum parameter set in future studies.

### 6.2. What Is a “Best” Model?

[38] As seen above, different metrics can provide different answers regarding model skill. It is therefore of interest to see whether the trends exhibited by the different misfit functions are similar or opposed to each other, and whether different metrics select for very different “best” model configurations. In particular, it is of interest for studies of biogeochemical tracer fluxes, whether we can identify any single model configuration that fits both observed phosphate and oxygen at the same time. In Figure 5 we have plotted the normalized misfit to surface phosphate (Figure 5, left) and total oxygen misfit (Figure 5, right) versus the total phosphate misfit for the different model experiments. Normalization was carried out by dividing each global misfit by the respective standard deviation of the observed tracer distribution.

[39] The “worst case” scenario when looking at two different metrics would correspond to an increase in the first misfit coinciding with a decrease in the second one, i.e. a line from the upper left corner to the lower right corner of Figure 5. Instead, the results indicate a positive correlation between the two metrics suggesting that there is at least the potential for simultaneously reducing both misfits. However, the relation between the two metrics depends on model type: while for the structurally simple models a large reduction especially in the surface misfit can be obtained for relatively small reductions in total phosphate misfit, the more complex model displays a large reduction in the total misfit with only a smaller reduction in surface misfit. Similarly, the simpler models may be improved more with respect to oxygen than



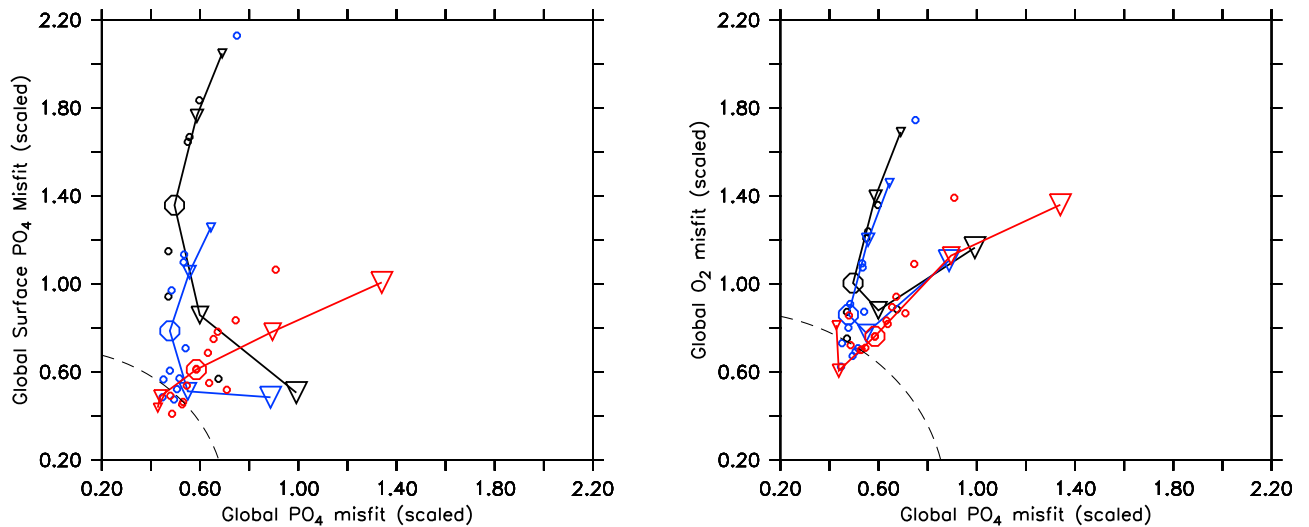


**Figure 4.** Global misfit of model NPZD-DOP for (left) phosphate and (right) oxygen for different sinking parameters  $r/a$ , and two different half-saturation constants for light. Thin line with open circles:  $I_c = 24 \text{ W m}^{-2}$  (default). Thick line with triangles:  $I_c = 48 \text{ W m}^{-2}$ . The symbols indicate the different experiments. See text for further details.

with respect to phosphate, while the complex model is about as sensitive to phosphate as it is to oxygen.

[40] To see if there is a specific class of model types and experiments that is optimal for two metrics at the same time, we have calculated the euclidean norm for each pair of misfits:  $E = \sqrt{m_1^2 + m_2^2}$  where  $m_1$  and  $m_2$  are either (normalized) surface and total phosphate misfit, or (normalized) total oxygen and total phosphate misfit. (In terms of Figure 5, this measure gives the distance of each data point from the origin.)

[41] Table 4 shows the ranking of the best eight models of this joint metric, for both combinations. For the combination of surface and total phosphate misfit the more complex model NPZD-DOP seems to be more appropriate than the simpler N-DOP model. Further, a decrease of production and/or export to deep layer in NPZD-DOP improves the combined fit. For the combination of oxygen and phosphate misfit, N-DOP appears more often than the more complex NPZD-DOP among the best eight candidates. For this simple model, an enhancement of export production either via a decrease in  $\sigma_{DOP}$  (the fraction of production released



**Figure 5.** (left) Surface layer misfit versus total misfit for phosphate and (right) total oxygen misfit versus total phosphate misfit for the different models. Misfits have been normalized by the respective standard deviation of observations (i.e., over  $x$  and  $y$  for the surface layer misfit, and over  $x$ ,  $y$  and  $z$  for the total misfits). The quarter circle denotes the max. euclidean norm of the best eight models shown in Table 4. Colors and symbols as in Figure 3.

**Table 4.** Model Ranking for the Euclidean Norm of Surface and Total Phosphate Misfit and Total Oxygen and Phosphate Misfit<sup>a</sup>

Rank	Surface and Total Phosphate			Total Oxygen and Total Phosphate		
	Model	Experiment	Description	Model	Experiment	Description
1	C	<i>s1</i>	very slow sinking	C	<i>s2</i>	slow sinking
2	C	<i>p4</i>	low light sensitivity	S	<i>p3</i>	high light sensitivity
3	C	<i>s2</i>	slow sinking	S	<i>d3</i>	high export ratio
4	S	<i>p3</i>	high light sensitivity	S	<i>d4</i>	high export ratio
5	S	<i>d3</i>	high export ratio	S	<i>p5</i>	high nutrient sensitivity
6	C	<i>p1</i>	slow growth rate	C	<i>p4</i>	low light sensitivity
7	C	<i>z2</i>	high grazing rate	S	<i>d5</i>	high export ratio
8	C	<i>z3</i>	high food sensitivity	C	<i>z2</i>	high grazing rate

<sup>a</sup>Model type is indicated by letter: “C” for more complex model NPZD + DOP, “S” for simpler model N+DOP. Model experiment is indicated by letter + number as in Tables 2 and 3.

as DOP), or via increase of nutrient or light sensitivity is beneficial. The more complex model, in contrast, benefits from a decrease in production parameters or remineralization length scale.

[42] Altogether, a simultaneous match to observed oxygen and phosphate distributions can be achieved with both structurally simple and complex models, but a simultaneous match to surface and total phosphate can be better accomplished with the more complex NPZD-DOP model. Fortunately, the same or similar candidates of both model types appear among the best candidates for both combined metrics. Configuration *s2* of the NPZD-DOP model with a slightly reduced remineralization length scale seems to be the most likely candidate to match the different criteria at the same time.

### 6.3. Biogeochemical Fluxes as Model Constraints?

[43] As shown above, the models’ global and regional fit to observed phosphate and oxygen distributions is most sensitive to parameters that regulate particle flux and/or production of organic matter. Because primary production, export production and remineralization profiles of organic matter are key controls on simulated biogeochemical tracer distributions, such model fluxes have often been compared against observational flux estimates [e.g., *Bacastow and Maier-Reimer*, 1991; *Najjar et al.*, 2007]. In the following, we investigate whether, and to what extent, such observational flux estimates can provide additional constraints on the model’s ability to reproduce observed biogeochemical tracer distributions.

[44] For primary production (Figures 6a–6c) we use observational estimates derived from remote sensing [*Carr et al.*, 2006]. To convert the fluxes (given in  $\text{Gt C y}^{-1}$ ) to global average values of  $\text{mmol P m}^{-2} \text{y}^{-1}$ , we use a molar C:P ratio of 106, and divide the global fluxes by the corresponding model area. (In the fixed-stoichiometry models, the model equivalent of primary production is phosphate uptake via phytoplankton.) We further compare our simulated primary production to the estimates by *Honjo et al.* [2008], which refer to regions with a water depth  $\geq 2000$  m.

[45] Most of the models predict lower than observed global primary production (Figures 6a–6c). According to our results, primary production is a poor predictor of model performance with respect to the misfit for phosphate, as a good fit to observed phosphate can be achieved both with high and low primary production (Figure 6b). Even a

presumably “correct” fit to observed primary production (as indicated by the vertical lines or shaded area) does not guarantee a good model fit to phosphate. The relation between global primary production and either surface phosphate misfit or oxygen misfit shows a similar pattern. Only model NPZD-DOP seems to be capable of reproducing the observed tracer distribution and simulating realistically high levels of primary production at the same time (Figures 6a and 6c).

[46] It remains to be investigated whether the simple models can be further improved with respect to observed primary production, e.g., via an increase in light or nutrient affinity, and/or modifications in the recycling parameters. Our results indicate that, while keeping a good fit to the observed phosphate distribution, we might be able to tune NPZD-DOP to match any desired primary production (for example, by parameterizing a “fast recycling loop” as in *Oschlies* [2001]). On the other hand, a good fit to observed primary production does not automatically imply a good fit to global phosphate or oxygen, as indicated by a few experiments of NPZD-DOP that lie within the range of observed primary production, but have an RMS misfit  $> 0.4 \text{ mmol P m}^{-3}$ . In case of surface phosphate, experiments of NPZD-DOP that fall within the range of observational primary-production estimates also show a small misfit.

[47] The range of observational estimates of export production (taken to be approximately the same as new production) as compiled by *Oschlies* [2001] is indicated in Figures 6d–6f, together with the observational estimate by *Lutz et al.* [2007] and *Honjo et al.* [2008]. Again we convert the fluxes via a molar C:P ratio of 106, and divide the global fluxes by the corresponding model area (area with water depth  $\geq 120$  m). As model equivalent for export production we use the particle flux at a depth of 120 m (i.e., out of the euphotic zone).

[48] Rates of global export production simulated by the different model configurations are confined to a narrower range than primary production and mostly fall between the values estimated by *Honjo et al.* [2008] and [from calibrated sediment traps *Lutz et al.* [2007], but are much lower than many of the estimates compiled by *Oschlies* [2001]. The review of estimated and simulated global export production by *Oschlies* [2001] suggests a wide range of potential values, part of which may be attributed to the effects of diapycnal diffusion and advection numerics in the different model types. Nevertheless, even with a “correct” simulation of export production a model could still fail to

represent the observed phosphate distribution (Figure 6e), as the values scatter quite strongly in the range between 10 and 15  $\text{mmol P m}^{-2} \text{y}^{-1}$ . Similar to the misfit to the observed global phosphate distribution, both the global misfit to oxygen and the surface misfit show little dependence on export production (Figures 6d and 6f).

[49] The simulated meso- and bathypelagic particle flux is compared to the observational estimates by *Honjo et al.* [2008] and *Lutz et al.* [2007]. The observed particle flux by *Honjo et al.* [2008] was extrapolated to the simulated depth of 2250 m by using a power law algorithm with an exponent of 0.86, as in their original data set. Again, we use

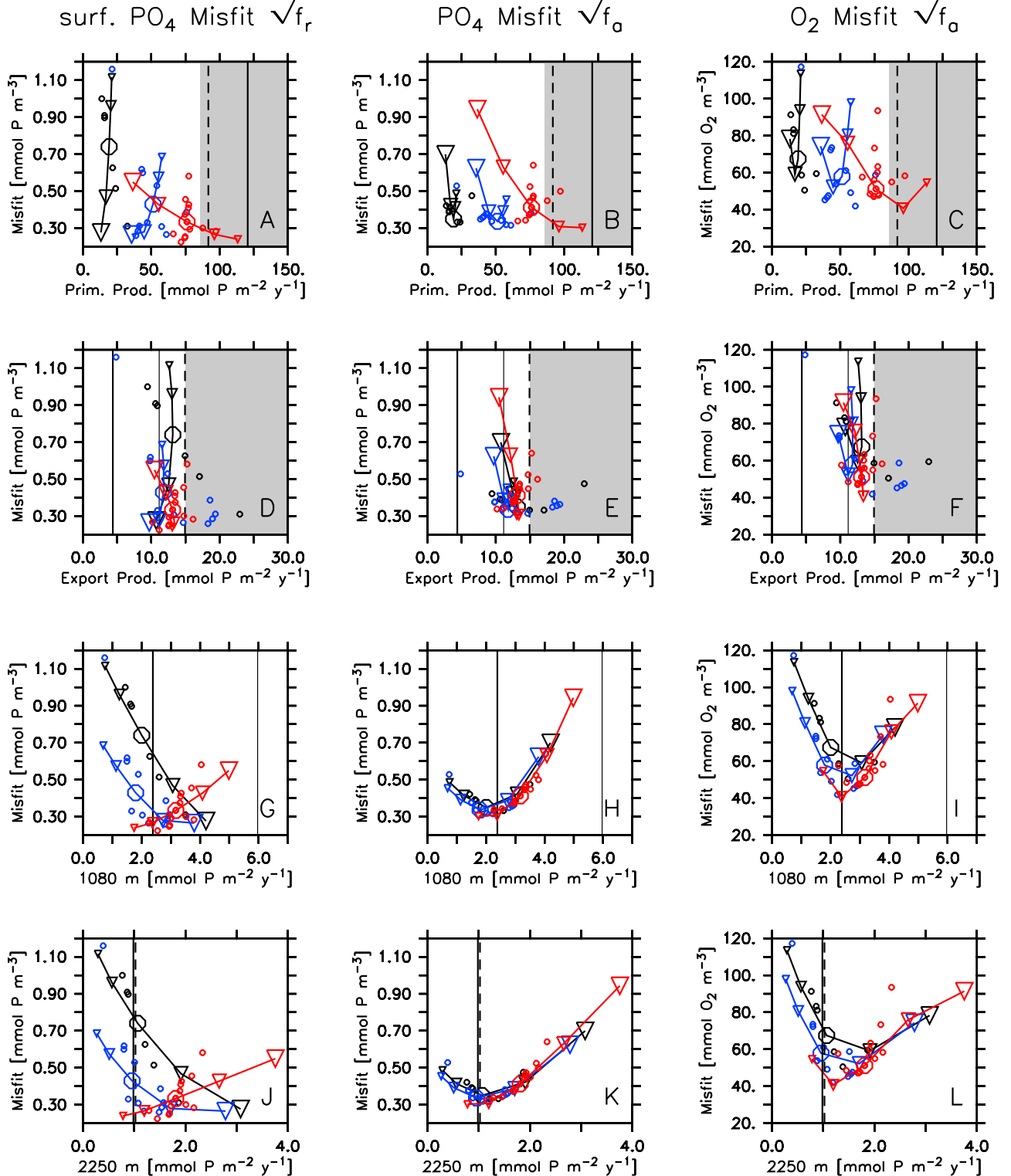


Figure 6

a C:P ratio of 106 to convert their carbon-based results to phosphorus. Although conceptually different from their work, a power law exponent of 0.86 has also been used for extrapolating the estimates by *Lutz et al.* [2007] to the simulated depths.

[50] Contrary to the results for primary production and export production, the misfit function for phosphate shows a clear parabolic dependence on simulated particle flux at 1080 and 2250 m depth (Figures 6h and 6k, respectively). Models with very high or very low deep particle flux perform poorly in terms of reproducing the observed phosphate distribution, and the misfit function shows a minimum RMS misfit for deep particle flux around  $2 \text{ mmol P m}^{-2} \text{ y}^{-1}$  and  $1 \text{ mmol P m}^{-2} \text{ y}^{-1}$  for particle flux at 1080 and 2250 m, respectively. In contrast to the near-surface fluxes associated with primary production and export production, this minimum RMS misfit agrees well with the observational estimates of the deep particle flux by *Lutz et al.* [2007], uncalibrated traps and *Honjo et al.* [2008]. At 1080 m it is far lower than the estimated from traps with radiogenic calibration, but the trapping efficiency from this calibration is still uncertain [*Lutz et al.*, 2007] and may vary with location and/or depth [*Scholten et al.*, 2001; *Yu et al.*, 2001].

[51] As for phosphate, the relation of the model misfit to observed oxygen distribution versus deep particle flux shows a rather clear dependency on the strength of simulated mesopelagic and bathypelagic particle flux (Figures 6i and 6l). Again, the global particle flux of the optimum experiments of each model is not very different from the observed deep particle flux.

[52] To summarize, in our experiments, primary production does not provide much information about the skill of models with respect to phosphate or oxygen distributions, as a good fit may be achieved with a variety of simulated primary production rates. On the other hand, only the more complex model NPZD-DOP is able to fit both production and phosphate at the same time. Global primary production might therefore serve as an additional constraint, especially if we want to represent biogeochemical fluxes in the surface layer. A realistic simulation of export production may still be associated with a relatively poor performance in terms of the simulated phosphate distribution. Particle fluxes in the meso- and bathypelagic realm, in contrast, seem to be a much better predictor of the models' fit to either phosphate or oxygen: Model solutions with a good fit to observed phosphate agree well with independent estimates of deep particle flux.

[53] However, care must be taken when comparing the global models to the flux estimates derived from observations, as these estimates are also based on certain assumptions about organism physiology and ecosystem structure,

which are not necessarily consistent with the assumptions made by the model. In that sense, the above comparison to observed fluxes has, to some extent, to be taken as a model-model intercomparison rather than a comparison to in-situ observations.

[54] In-situ observations of biogeochemical fluxes are rather sparse in space and time, and the physical context in which these were obtained may differ strongly from the usually coarsely-resolved physics in the global models presented here. Further, the methodology of these observations is often not as standardized as for the chemical tracers. We therefore regard the information gained from observational flux estimates as somewhat weaker constraint than the information gained from biogeochemical tracer measurements. In this respect, the comparison to global biogeochemical fluxes can be viewed as an additional aid in model development that may supplement the quantitative model assessment with respect to the distributions of tracers such as nutrients or oxygen.

## 7. Summary and Conclusions

[55] A number of structurally different biogeochemical models have been assessed for a range of different parameter settings against observed distributions of phosphate and oxygen. Besides illustrating that such a quantitative model assessment is technically feasible, the aim of this study was to examine the sensitivity of the model-data misfits to variations in parameter values and to variations in biogeochemical model structure. Our analysis suggests that the parameterization of particle flux or sinking of detritus plays a large role for a realistic distribution of tracers among the different oceanic regions, in particular between the northern North Atlantic and northern North Pacific. While these two regions contribute very little to the global misfit, especially the northern North Pacific's misfit, in particular, can be quite large. Therefore, care must be taken when attempting to derive information about particular regions from models that have been designed to give a good global fit.

[56] While all models show roughly the same sensitivity in their fit to global phosphate to variations in model parameters, they differ strongly with respect to the impact of parameter variations on nutrient misfit derived only for the surface layers. The structurally more complex NPZD-DOP model is less sensitive to variations in single parameters than are the simpler models. Whether this is due to the different starting points in the parameter space for the different model types (as indicated by their different biological timescales), or due to a higher complexity of the NPZD-DOP model remains to be investigated in a more comprehensive search of the parameter space. For the parameter combinations

**Figure 6.** Misfit functions versus different global biogeochemical fluxes. (a, d, g, and j) Surface misfit of phosphate. (b, e, h, and k) Total misfit of phosphate. (c, f, i, and l) Total misfit of oxygen. Figures 6a–6c show misfit versus primary production. Figures 6d–6f show misfit versus export production. Figures 6g–6i show misfit versus particle flux in 1080 m. Figures 6j–6l show misfit versus particle flux in 2250 m. Colors, symbols and lines denote different model experiments, as in Figure 3. Vertical lines: observed fluxes. Primary production (Figures 6a–6c): *Carr et al.* [2006] (average over 24 different algorithms, straight line) and *Honjo et al.* [2008] (dashed line). Export production (Figures 6d–6f): *Lutz et al.* [2007] (thick straight: uncalibrated, thin straight: radiogenic calibration) and *Honjo et al.* [2008] (dashed). Particle flux in 1080 m: *Lutz et al.* [2007] (thick straight: uncalibrated, thin straight: radiogenic calibration). Particle flux in 2250 m: *Lutz et al.* [2007] (thick straight: uncalibrated) and *Honjo et al.* [2008] (dashed). Shaded area in the plots for primary and export production denote the range given by *Carr et al.* [2006, Figure 5A] and *Oschlies* [2001], respectively.

tested in this study, it appears that structurally more complex models may be better able to reproduce observed surface nutrient distributions, whereas structurally simple and structurally complex models seem to perform similarly well in reproducing phosphate and oxygen distributions in the global ocean interior.

[57] Overall, various metrics and model assessments of our (limited) evaluation of model types and setups suggest that parameters involved in the description of the particle flux are among the most important ones in describing global ocean distributions of nutrients and oxygen. For the model configurations used here, the corresponding exponent of the Martin function used to describe particle flux [Martin *et al.*, 1987] lies within the range of  $\approx 1.1$ – $0.86$ . This agrees with other model studies [e.g., Kwon and Primeau, 2006], and observational estimates [e.g., Martin *et al.*, 1987]. Our model comparison also supports the global deep flux estimates by Honjo *et al.* [2008] and Lutz *et al.* [2007]. It remains to be investigated whether this result is confirmed for biogeochemical models simulated with different spatial resolution and different circulation fields. Sensitivity studies with transport matrices derived from a higher-resolution ( $1^\circ \times 1^\circ$  laterally and 23 levels vertically) global model found smallest phosphate and oxygen misfits for virtually the same export parameters.

[58] While sensitivity studies of zero-dimensional or one-dimensional oceanic biogeochemical models have long been carried out [e.g., Fasham *et al.*, 1990], this has generally not been the case for computationally much more expensive three-dimensional global biogeochemical models. To our knowledge, this study presents the first attempt to systematically examine the parameter sensitivity of a range of global biogeochemical models. Our study suggest two broad avenues for future research. First, it is important to investigate how robust our results are to the spatial resolution and circulation of the underlying physical model. Here, for computational efficiency, we have restricted our attention to a relatively coarse resolution ocean model. Further work is needed to better characterize and understand this sensitivity to the physical circulation. Second, proceeding along the lines of Fasham *et al.* [1990], it would be useful to optimize of the biogeochemical models, as carried out in a zero-dimensional context by, for example, Fasham and Evans [1995] or Schartau and Oschlies [2003a, 2003b]. Optimization and sensitivity analysis of global biogeochemical models - such as carried out by Kwon and Primeau [2006, 2008] for a structurally simple model - can provide more insight into the dynamics of a global modeling framework (i.e., biogeochemistry + ocean circulation). However, for more complex models, such as the NPZD-DOP one presented in this study, this may be a demanding task. Reducing the parameter set to be optimized could help to make this task more feasible. Our results suggest that the particle-flux and remineralization parameters are likely to emerge as particularly important in terms of a model's ability to reproduce global biogeochemical tracer patterns, in the ocean interior. In the current study we could not identify an improvement in the simulation of ocean-interior nutrient and oxygen distributions with increased structural model complexity. The more complex model used here was, however, able to fit observed surface nutrient distributions better than

the structurally simpler models. The degree of model complexity required may thus depend on the metric used for evaluation which, in turn, should reflect the scientific question to be addressed by the model.

## Appendix A: Biogeochemical Model Description

[59] We assume that different biogeochemical processes operate in different vertical domains. In the upper layers of the ocean a fast and dynamic turnover of phosphorus via photosynthesis, grazing, mortality, excretion and/or exudation takes place. In the aphotic layers there is only a slow turnover of phosphorus. To specify processes operating only in the euphotic zone ( $0$ – $120$  m, or  $k \leq k_e = 2$ ), we use the symbol  $\mathcal{H}_e(k) \equiv H(k_e - k)$ , where  $H(k)$  is the Heaviside step function.  $\mathcal{H}_e(k)$  is “1” in the euphotic layers, and “0” outside, conveniently allowing us to set processes outside the euphotic layers to zero.

[60] All models presented here include oxygen. The air-sea gas exchange (top layer only) is parameterized following the OCMIP-2 protocol [Najjar and Orr, 1999], with piston velocity and saturation computed from a monthly mean wind speed, temperature and salinity derived from the MIT ocean model, and interpolated linearly onto the current time step. In all layers, oxygen also changes due to photosynthesis and remineralization, except in the presence of suboxic conditions (defined here as  $O_2 \leq 4$  mmol  $O_2$   $m^{-3}$ ). In this case oxygen does not change due to the biogeochemical processes:

$$S(O_2) = \begin{cases} -R_{O_2:P} S(PO_4) & : O_2 > 4 \text{ mmol } m^{-3} \\ 0 & : O_2 \leq 4 \text{ mmol } m^{-3} \end{cases} \quad (A1)$$

where  $R_{O_2:P}$  is the molar ratio  $O_2:P$  (chosen as 170) that relates sources and sinks of  $PO_4$  to those of  $O_2$ .

### A1. Models N and N-DOP

[61] In the euphotic layers, the simpler models calculate nutrient uptake by phytoplankton of an “implicit” phytoplankton concentration,  $PHY^* = const.$ . Primary production  $PP$  is calculated via multiplicative limitation by light and nutrients, i.e.  $PP = \mu_{PHY} PHY^* f(I) [PO_4/(K_{PHY} + PO_4)]$ .  $\mu_{PHY}$  is the maximum growth rate of the (implicitly prescribed) phytoplankton and  $K_{PHY}$  the half-saturation constant for nutrient uptake. Light limitation is parameterized via a Monod function:  $f(I) = I/(I_c + I)$ .  $I$  is 24-hour mean PAR (photosynthetically available radiation) at the center of each (euphotic) layer, taking into account the light attenuation by water and dissolved substances ( $k_w = 0.02$   $m^{-1}$ ). All uptake of phosphate is shifted either to export production or DOP. The partitioning between neutrally buoyant DOP and export  $E$  is regulated via the parameter  $\sigma_{DOP}$ . Export out of each source (euphotic) layer with index  $s = 1, 2$  is parameterized via  $E(s) = (1 - \sigma_{DOP}) PP(s) \Delta z(s)$ . The export flux  $E(s)$  out of each box is distributed over the layers below according to a power law.

$$F(k) = \sum_{s=1}^{s=2} E(s) \left( \frac{z(k)}{z(s+1)} \right)^{-r/a} \quad \text{for } k > 2, \quad (A2)$$

[62]  $z(s + 1)$  denotes the lower boundary of a source (exporting) layer, and  $z(k)$  the upper boundary of a receiving layer. The parameter  $r$  represents an implicit, constant remineralization rate of sinking organic matter, while  $a$  prescribes the rate of increase of sinking speed with depth (see equation (1)). Flux divergence directly feeds into the phosphate pool; thus phosphate is reduced by phytoplankton production in the euphotic zone, and increases due to the flux divergence and DOP decay in the entire water column. DOP in all layers remineralizes with a constant rate  $\lambda'_{DOP}$ , but only down to a lower limit  $P_{\min} = 10^{-6} \text{ mmol P m}^{-3}$ . The source-minus-sink terms for model “N-DOP” are therefore:

$$S(PO4) = -PP\mathcal{H}_e(k) + \frac{\partial F}{\partial z} + \lambda'_{DOP} \max(0, DOP - P_{\min}). \quad (\text{A3})$$

$$S(DOP) = \sigma_{DOP} PP\mathcal{H}_e(k) - \lambda'_{DOP} \max(0, DOP - P_{\min}). \quad (\text{A4})$$

[63] The “N” model considers only inorganic nutrients, and not the slowly decaying DOP, which can be parameterized by setting  $\sigma_{DOP} = 0$  and  $\lambda'_{DOP} = 0$  in (equations A3) and (A4).

## A2. Model NPZD-DOP

[64] This model simulates phytoplankton, zooplankton and detritus, the latter with a sinking speed increasing linearly with depth. It further differs to the simpler models N and N-DOP in several functional aspects: (1) Attenuation of light depends on water ( $k_w = 0.04 \text{ m}^{-1}$ ) and phytoplankton concentration ( $k_c = 0.48 \text{ m}^{-2} (\text{mmol P})^{-1}$ ). (2) Phytoplankton light limitation is parameterized following *Evans and Parslow* [1985]. This approach differs from the function used for N and N-DOP in essentially two ways: (a) The functional response of phytoplankton to light follows the function of *Smith* [1936], and (b) the light profile within a grid box is used to compute average phytoplankton growth over the entire layer. (3) We parameterize temperature dependent growth of phytoplankton by  $\mu_{PHY}(T) = \mu_{PHY} e^{\frac{T-T_B}{T_B}}$ , with  $T_B = 15.65^\circ\text{C}$  (as *Eppley* [1972], in the notation by *Schmittner et al.* [2008]). (4) While models N and N-DOP assume a multiplicative resource limitation, we now assume that the most limiting resource sets the phytoplankton growth rate, i.e.,  $PP = \mu_{PHY} \min(f(I), PO4/(K_{PHY} + PO4))$ .

[65] Grazing by zooplankton is described by a Holling-III function, i.e. via a quadratic dependence on phytoplankton, a maximum grazing rate  $\mu_{ZOO}$  and half-saturation constant  $K_{ZOO}$  of zooplankton. We assume that a fraction  $\sigma_{DOP}$  of egestion, zooplankton mortality, and phytoplankton loss is released as DOP, and the rest as detritus. DOP in all layers remineralizes with a constant rate  $\lambda'_{DOP}$ , but only above a lower limit  $P_{\min} = 10^{-6} \text{ mmol P m}^{-3}$ . Phytoplankton and zooplankton die with a constant mortality rate of  $\lambda'_{PHY} = \lambda'_{ZOO} = 0.01 \text{ d}^{-1}$ , again above a lower concentration threshold  $P_{\min}$ . We assume that the sinking speed of detritus is a property of detritus. Because detritus is defined at the center of each vertical box, we calculate its sinking speed (equation (1)) from the depth of the box centers. It

remineralizes with a fixed rate  $\lambda'_{DET} = 0.05 \text{ d}^{-1}$  directly to phosphate. Thus, the source-minus-sink terms are:

$$S(PO4) = (-PP + \lambda_{ZOO} ZOO)\mathcal{H}_e(k) + \lambda'_{DOP} \max(0, DOP - P_{\min}) + \lambda'_{DET} DET^* \quad (\text{A5})$$

$$S(DOP) = \sigma_{DOP} [(1 - \epsilon_{ZOO})G + \kappa_{ZOO} ZOO^2 + \lambda_{PHY} PHY]\mathcal{H}_e(k) - \lambda'_{DOP} \max(0, DOP - P_{\min}) + \lambda_{PHY} \max(0, PHY - P_{\min}) + \lambda'_{ZOO} \max(0, ZOO - P_{\min}) \quad (\text{A6})$$

$$S(PHY) = (PP - G - \lambda_P PHY)\mathcal{H}_e(k) - \lambda_{PHY} \max(0, PHY - P_{\min}) \quad (\text{A7})$$

$$S(ZOO) = (\epsilon_{ZOO} G - \lambda_{ZOO} ZOO - \kappa_{ZOO} ZOO^2)\mathcal{H}_e(k) - \lambda'_{ZOO} \max(0, ZOO - P_{\min}), \quad (\text{A8})$$

$$S(DET) = (1 - \sigma_{DOP}) [(1 - \epsilon_{ZOO})G + \kappa_{ZOO} ZOO^2 + \lambda_{PHY} PHY] \cdot \mathcal{H}_e(k) - \lambda'_{DET} DET^* + \frac{\partial w DET^*}{\partial z},$$

with  $DET^* = \max(0, DET - DET_{\min}) \quad (\text{A9})$

[66] **Acknowledgments.** This work is a contribution of the DFG-supported project SFB754. SK was funded by U.S. NSF grant OCE 08-24635. We thank two anonymous reviewers for their constructive and helpful comments. LDEO contribution 7537.

## References

- Aumont, O., and L. Bopp (2006), Globalizing results from ocean in situ iron fertilization studies, *Global Biogeochem. Cycles*, 20, GB2017, doi:10.1029/2005GB002591.
- Bacastow, R., and E. Maier-Reimer (1990), Ocean-circulation model of the carbon cycle, *Clim. Dyn.*, 4, 95–125.
- Bacastow, R., and E. Maier-Reimer (1991), Dissolved organic carbon in modeling oceanic new production, *Global Biogeochem. Cycles*, 5(1), 71–85.
- Buesseler, K., et al. (2007), Revisiting carbon flux through the ocean’s twilight zone, *Science*, 316, 567–570, doi:10.1126/science.1137959.
- Carr, M.-E., et al. (2006), A comparison of global estimates of marine primary production from ocean color, *Deep Sea Res., Part II*, 53, 741–770, doi:10.1016/j.dsr2.2006.01.028.
- Eppley, R. W. (1972), Temperature and phytoplankton growth in the sea, *Fish. Bull.*, 70, 1063–1085.
- Evans, G. T., and J. S. Parslow (1985), A model of annual plankton cycles, *Biol. Oceanogr.*, 3, 327–347.
- Fasham, M., and G. Evans (1995), The use of optimization techniques to model marine ecosystem dynamics at the JGOFS station at 47°N, 20°W, *Philos. Trans. R. Soc. B*, 348, 203–209.
- Fasham, M., H. Ducklow, and S. McKelvie (1990), A nitrogen-based model of plankton dynamics for the oceanic mixed layer, *J. Mar. Res.*, 48, 591–639.
- Garcia, H. E., R. A. Locarnini, T. P. Boyer, and J. I. Antonov (2006a), *World Ocean Atlas 2005*, vol. 3, *Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation*, NOAA Atlas NESDIS, vol. 63, edited by S. Levitus, U.S. Gov. Print. Off., Washington, D. C.
- Garcia, H. E., R. A. Locarnini, T. P. Boyer, and J. I. Antonov (2006b), *World Ocean Atlas 2005*, vol. 4, *Nutrients (Phosphate, Nitrate, Silicate)*, NOAA Atlas NESDIS, vol. 64, edited by S. Levitus, U.S. Gov. Print. Off., Washington, D. C.
- Honjo, S., S. Manganini, R. Krishfield, and R. Francois (2008), Particulate organic carbon fluxes to the ocean interior and factors controlling the biological pump: A synthesis of global sediment trap programs since 1983, *Prog. Oceanogr.*, 76, 217–285, doi:10.1016/j.pocean.2007.11.003.



- Khatiwala, S. (2007), A computational framework for simulation of biogeochemical tracers in the ocean, *Global Biogeochem. Cycles*, 21, GB3001, doi:10.1029/2007GB002923.
- Khatiwala, S. (2008), Fast spin up of ocean biogeochemical models using matrix-free Newton-Krylov, *Ocean Modell.*, 23, 121–129, doi:10.1016/j.ocemod.2008.05.002.
- Khatiwala, S., M. Visbeck, and M. A. Cane (2005), Accelerated simulation of passive tracers in ocean circulation models, *Ocean Modell.*, 9, 51–69.
- Khatiwala, S., F. Primeau, and M. Holzer (2012), Ventilation of the deep ocean constrained with tracer observations and implications for radio-carbon estimates of ideal mean age, *Earth Planet. Sci. Lett.*, 325–326, 116–125.
- Kriest, I., and A. Oschlies (2008), On the treatment of particulate organic matter sinking in large-scale models of marine biogeochemical cycles, *Biogeosciences*, 5, 55–72.
- Kriest, I., and A. Oschlies (2011), Numerical effects on organic-matter sedimentation and remineralization in biogeochemical ocean models, *Ocean Modell.*, 39, 275–283, doi:10.1016/j.oceanmod.2011.05.001.
- Kriest, I., S. Khatiwala, and A. Oschlies (2010), Towards an assessment of simple global marine biogeochemical models of different complexity, *Prog. Oceanogr.*, 86, 337–360, doi:10.1016/j.pocean.2010.05.002.
- Kwon, E. Y., and F. Primeau (2006), Optimization and sensitivity study of a biogeochemistry ocean model using an implicit solver and in situ phosphate data, *Global Biogeochem. Cycles*, 20, GB4009, doi:10.1029/2005GB002631.
- Kwon, E. Y., and F. Primeau (2008), Optimization and sensitivity of a global biogeochemistry ocean model using combined in situ DIC, alkalinity, and phosphate data, *J. Geophys. Res.*, 113, C08011, doi:10.1029/2007JC004520.
- Kwon, E. Y., F. Primeau, and J. Sarmiento (2009), The impact of remineralization depth on the air-sea carbon balance, *Nat. Geosci.*, 2, 630–635, doi:10.1038/NGEO612.
- Li, X., and F. Primeau (2008), A fast Newton-Krylov solver for seasonally varying global ocean biogeochemistry models, *Ocean Modell.*, 23(1–2), 13–20, doi:10.1016/j.ocemod.2008.03.001.
- Lutz, M. J., K. Caldeira, R. B. Dunbar, and M. J. Behrenfeld (2007), Seasonal rhythms of net primary production and particulate organic carbon flux to depth describe the efficiency of biological pump in the global ocean, *J. Geophys. Res.*, 112, C10011, doi:10.1029/2006JC003706.
- Marshall, J., A. Adcroft, C. Hill, L. Perelman, and C. Heisey (1997), A finite-volume, incompressible Navier-Stokes model for studies of the ocean on parallel computers, *J. Geophys. Res.*, 102, 5733–5752.
- Martin, J. H., G. A. Knauer, D. M. Karl, and W. W. Broenkow (1987), VERTEX: Carbon cycling in the northeast Pacific, *Deep Sea Res.*, 34(2), 267–285.
- Moore, J. K., and S. C. Doney (2007), Iron availability limits the ocean nitrogen inventory stabilizing feedbacks between marine denitrification and nitrogen fixation, *Global Biogeochem. Cycles*, 21, GB2001, doi:10.1029/2006GB002762.
- Najjar, R., and J. C. Orr (1999), Biotic-HOWTO, internal OCMIP report, Lab. des Sci. du Clim. et l'Environ., CEA Saclay, Gif-sur-Yvette, France.
- Najjar, R. G., et al. (2007), Impact of circulation on export production, dissolved organic matter and dissolved oxygen in the ocean: Results from Phase II of the Ocean Carbon-cycle Model Intercomparison Project (OCMIP-2), *Global Biogeochem. Cycles*, 21, GB3007, doi:10.1029/2006GB002857.
- Oschlies, A. (2001), Model-derived estimates of new production: Why higher resolution may imply lower values, *Deep Sea Res. Part II*, 48, 2173–2197.
- Parekh, P., M. J. Follows, and E. A. Boyle (2005), Decoupling of iron and phosphate in the global ocean, *Global Biogeochem. Cycles*, 19, GB2020, doi:10.1029/2004GB002280.
- Schartau, M., and A. Oschlies (2003a), Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part I—Method and parameter estimates, *J. Mar. Res.*, 61, 765–793.
- Schartau, M., and A. Oschlies (2003b), Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part II—Standing stocks and nitrogen fluxes, *J. Mar. Res.*, 61, 795–821.
- Schmittner, A., A. Oschlies, X. Giraud, M. Eby, and H. L. Simmons (2005), A global model of the marine ecosystem for long-term simulations: Sensitivity to ocean mixing, buoyancy forcing, particle sinking, and dissolved organic matter cycling, *Global Biogeochem. Cycles*, 19, GB3004, doi:10.1029/2004GB002283.
- Schmittner, A., A. Oschlies, H. D. Matthews, and E. D. Galbraith (2008), Future changes in climate, ocean circulation, ecosystems, and biogeochemical cycling simulated for a business-as-usual CO<sub>2</sub> emission scenario until year 4000 AD, *Global Biogeochem. Cycles*, 22, GB1013, doi:10.1029/2007GB002953.
- Scholten, J., J. Fietzke, S. Vogler, M. Rutgers van der Loeff, A. Mangini, W. Koeve, J. Wanek, P. Stoffers, A. Antia, and J. Kuss (2001), Trapping efficiencies of sediment traps from the deep eastern North Atlantic: The <sup>230</sup>Th calibration, *Deep Sea Res. Part II*, 48, 2383–2408.
- Smith, E. L. (1936), Photosynthesis in relation to light and carbon dioxide, *Proc. Natl. Acad. Sci. U. S. A.*, 22, 504–511.
- Yu, E.-F., R. Francois, M. P. Bacon, S. Honjo, A. P. Fleer, S. J. Manganini, M. M. R. van der Loeff, and V. Ittekkot (2001), Trapping efficiency of bottom-tethered sediment traps estimated from the intercepted fluxes of <sup>230</sup>Th and <sup>231</sup>Pa, *Deep Sea Res.*, 48(3), 865–889.